Mai Surintraboon
9_6204640178

EE426 Final Exam

1) From the data set "Final_q1-1_no.dta":

(a) Estimate the model using multinomial logit model using yi=0 is the base outcome. Make interpretation of your estimated results (sign & meaning (in term of rrr), overall test, goodness of fit, and individual test).

```
. mlogit y x1 x2 x3 x4, base(0) nolog

Multinomial logistic regression                 Number of obs     =        170
                                                LR chi2(8)        =      90.21
                                                Prob > chi2       =     0.0000
Log likelihood = -102.72843                     Pseudo R2         =     0.3051

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
          x1 |  -.8808854   .6242729    -1.41   0.158    -2.104438    .3426669
          x2 |   -.547963   .7885212    -0.69   0.487    -2.093436    .9975101
          x3 |  -.4119661   .6294599    -0.65   0.513    -1.645685    .8217527
          x4 |   .5520053   .2635367     2.09   0.036     .0354828    1.068528
       _cons |  -6.536554   3.711431    -1.76   0.078    -13.81082    .7377173
-------------+----------------------------------------------------------------
2            |
          x1 |  -.6606928   .7073902    -0.93   0.350    -2.047152    .7257664
          x2 |  -1.785442   .8242509    -2.17   0.030    -3.400944   -.1699395
          x3 |   .6424234    .676681     0.95   0.342     -.683847    1.968694
          x4 |   1.828801   .3168299     5.77   0.000     1.207826    2.449776
       _cons |  -26.40387   4.699289    -5.62   0.000    -35.61431   -17.19344
------------------------------------------------------------------------------

. est store my

. fitstat, saving(my)

Measures of Fit for mlogit of y
```

Log-Lik Intercept Only:	-147.832	Log-Lik Full Model:	-102.728
D(155):	205.457	LR(8):	90.207
		Prob > LR:	0.000
McFadden's R2:	0.305	McFadden's Adj R2:	0.204
ML (Cox-Snell) R2:	0.412	Cragg-Uhler(Nagelkerke) R2:	0.500
Count R2:	0.759	Adj Count R2:	0.317
AIC:	1.385	AIC*n:	235.457
BIC:	-590.592	BIC':	-49.121
BIC used by Stata:	256.815	AIC used by Stata:	225.457

(Indices saved in matrix fs_my)

```
. mlogit y x1 x2 x3 x4, base(0) rrr nolog
```

Multinomial logistic regression		Number of obs	= 170
		LR chi2(8)	= 90.21
		Prob > chi2	= 0.0000
Log likelihood = -102.72843		Pseudo R2	= 0.3051

y	RRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]
0	(base outcome)					
1						
x1	.4144158	.2587085	-1.41	0.158	.1219142	1.408699
x2	.5781262	.4558648	-0.69	0.487	.1232629	2.711522
x3	.6623467	.4169207	-0.65	0.513	.1928804	2.274483
x4	1.736732	.4576927	2.09	0.036	1.03612	2.911091
_cons	.0014495	.0053796	-1.76	0.078	1.00e-06	2.091156
2						
x1	.5164934	.3653623	-0.93	0.350	.1291021	2.066314
x2	.167723	.1382458	-2.17	0.030	.0333418	.8437158
x3	1.901082	1.286426	0.95	0.342	.5046718	7.161316
x4	6.226418	1.972716	5.77	0.000	3.346202	11.58576
_cons	3.41e-12	1.60e-11	-5.62	0.000	3.41e-16	3.41e-08

The sign of the estimated results has to be check whether they follow the theory or not.

For the meaning in terms of RRR, our estimated result in both cases (y=1 and y=2) has a positive sign. This means that when the variable x increases, the probability of choosing that alternative instead of the baseline (y=0) increases.

For example, let's interpret x1's RRR: 0.4144158 in the y=1 case. It means that when x1 increases, the probability of choosing y=1 [P(y=1)] instead of the baseline y=0 [P(y=0)] increases.

From the LR test above as the p-value is 0.0000 < 0.05, we can reject the null hypothesis that the estimated results are overall insignificant. All the estimated results are overall significant.

The individual Z-test suggest that for y=1, x1, x2, x3 are insignificant at 95% confidence level as they presented the p-values more than 0.05. x4 is significant as its p-value is 0.036 < 0.05.

Moreover, for y=2, x1 and x3 are insignificant while x2 is significant at 95% confidence level.

The counted R2 of the model is 0.759 > 0.05. From this R2 result, it can be said that the model fit the data quite well.

Perform IIA test whether it is appropriated to apply multinomial logit in this case? Give explanation why IIA is important for multinomial logit. Determine whether multinomial logit is appropriated for this case? Why?

```
. *Perform IIA Test

. mlogit y x1 x2 x3 x4 if y!=1, base(0) nolog


Multinomial logistic regression                 Number of obs     =        129
                                                LR chi2(4)        =      54.40
                                                Prob > chi2       =     0.0000
Log likelihood = -26.719102                     Pseudo R2         =     0.5045

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
2            |
          x1 |  -1.165754   .9346013    -1.25   0.212    -2.997539    .6660305
          x2 |  -2.671481   1.107519    -2.41   0.016    -4.842178   -.5007836
          x3 |    .312561   .7716918     0.41   0.685    -1.199927    1.825049
          x4 |   2.107134   .4611287     4.57   0.000     1.203338    3.010929
       _cons |  -29.67513   6.579104    -4.51   0.000    -42.56993   -16.78032
------------------------------------------------------------------------------
```

```
. est store myno1

. hausman myno1 my, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     myno1          my         Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |   -1.165754    -.6606928       -.5050616        .6108017
          x2 |   -2.671481    -1.785442       -.8860391        .7397354
          x3 |     .312561     .6424234       -.3298624        .3709596
          x4 |    2.107134     1.828801        .2783326        .3350499
       _cons |   -29.67513    -26.40387       -3.271254        4.604486
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from mlogit
            B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        3.41
                Prob>chi2 =      0.6366
```

From the IIA hausman test, the p-value equals to 0.6366>0.05. Thus, IIA hypothesis is not rejected. IIA is satisfied in this model, meaning it is appropriate to use multinomial logit in this case.

IIA is important because it implies that the decision between 2 alternatives is independent from the existence of more alternatives.

(b) Estimate the model using order probit model of yi. Make interpretation of your estimated results (sign & meaning, overall test, goodness of fit, and individual test).

```
. oprobit y x1 x2 x3 x4, nolog

Ordered probit regression                       Number of obs     =        170
                                                LR chi2(4)        =      81.28
                                                Prob > chi2       =     0.0000
Log likelihood = -107.19118                     Pseudo R2         =     0.2749
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	-.1496765	.2384153	-0.63	0.530	-.6169619	.3176089
x2	-.5828778	.2454394	-2.37	0.018	-1.06393	-.1018254
x3	.4722078	.2223359	2.12	0.034	.0364374	.9079783
x4	.7043575	.0936777	7.52	0.000	.5207527	.8879624
/cut1	9.692317	1.418965			6.911197	12.47344
/cut2	10.94181	1.474225			8.052381	13.83124

```
. est store oy

. fitstat, saving(oy)

Measures of Fit for oprobit of y

Log-Lik Intercept Only:       -147.832   Log-Lik Full Model:          -107.191
D(164):                        214.382   LR(4):                        81.282
                                         Prob > LR:                     0.000
McFadden's R2:                   0.275   McFadden's Adj R2:             0.234
ML (Cox-Snell) R2:               0.380   Cragg-Uhler(Nagelkerke) R2:    0.461
McKelvey & Zavoina's R2:         0.497
Variance of y*:                  1.989   Variance of error:             1.000
Count R2:                        0.735   Adj Count R2:                  0.250
AIC:                             1.332   AIC*n:                       226.382
BIC:                          -627.889   BIC':                        -60.739
BIC used by Stata:             245.197   AIC used by Stata:           226.382

(Indices saved in matrix fs_oy)

. *Marginal effect for order probit

. est restore oy
(results oy are active now)
```

Marginal effect at mean

```
. mfx, predict(outcome(0))

Marginal effects after oprobit
      y  = Pr(y==0) (predict, outcome(0))
         =  .03868796
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0125529      .02024    0.62   0.535  -.027116  .052221   .476264
      x2*|   .0422275      .01887    2.24   0.025   .005234  .079221   .670588
      x3 |  -.0396025       .0209   -1.90   0.058  -.080556  .001351   .895059
      x4 |  -.0590722       .0163   -3.62   0.000  -.091013 -.027132    16.324
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx, predict(outcome(1))

Marginal effects after oprobit
      y  = Pr(y==1) (predict, outcome(1))
         =  .26402029
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0396995       .0636    0.62   0.532   -.08495  .164349   .476264
      x2*|    .148435      .06149    2.41   0.016   .027924  .268946   .670588
      x3 |  -.1252463      .06074   -2.06   0.039  -.244288 -.006205   .895059
      x4 |  -.1868207      .03562   -5.24   0.000   -.25664 -.117001    16.324
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx, predict(outcome(2))

Marginal effects after oprobit
      y  = Pr(y==2) (predict, outcome(2))
         =  .69729175
------------------------------------------------------------------------------
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	-.0522524	.08339	-0.63	0.531	-.215696	.111191	.476264
x2*	-.1906625	.07426	-2.57	0.010	-.336211	-.045114	.670588
x3	.1648488	.07702	2.14	0.032	.013886	.315811	.895059
x4	.2458928	.03357	7.33	0.000	.180105	.311681	16.324

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Marginal effect at median

```
. mfx, at(median) predict(outcome(0))

Marginal effects after oprobit
      y  = Pr(y==0) (predict, outcome(0))
         =  .03005679
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.0102003	.01703	0.60	0.549	-.023176	.043576	.7
x2*	.0231647	.01229	1.88	0.059	-.000922	.047251	1
x3	-.0321804	.01515	-2.12	0.034	-.061883	-.002478	1.175
x4	-.0480011	.01762	-2.72	0.006	-.082531	-.013471	16.6181

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
. mfx, at(median) predict(outcome(1))

Marginal effects after oprobit
      y  = Pr(y==1) (predict, outcome(1))
         =  .23413746
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.0387494	.06234	0.62	0.534	-.083442	.160941	.7
x2*	.1285307	.05214	2.47	0.014	.026344	.230718	1
x3	-.1222488	.0567	-2.16	0.031	-.233387	-.011111	1.175
x4	-.1823496	.03452	-5.28	0.000	-.250001	-.114698	16.6181

```
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1


. mfx, at(median) predict(outcome(2))


Marginal effects after oprobit
      y  = Pr(y==2) (predict, outcome(2))
         =  .73580575
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |  -.0489497      .07891   -0.62   0.535  -.203618  .105718        .7
      x2*|  -.1516954      .06005   -2.53   0.012  -.269396 -.033995         1
      x3 |   .1544292      .06653    2.32   0.020   .024038  .284821     1.175
      x4 |   .2303506      .03678    6.26   0.000   .158257  .302444   16.6181
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

While x1 and x2 possess a negative sign, x3 and x4 possess a positive sign. These signs should be checked their accordance with the theory.

For the meaning of this estimated result, we should look at the marginal effects. Shown above are the marginal effects both at the mean and median of all estimated results in every predicted outcome. I will illustrate the interpretation of one marginal effect at mean and one at median. Noted that the way other cases can be estimated is the same.

M.E. at mean interpretation for x2 at P(y=0): If x2 changes from its mean by 1 unit, the probability that y equals to 0 will increase by 4.22275%.

M.E. at median interpretation for x2 at P(y=1): If x2 changes from its median by 1 unit, the probability that y equals to 1 will increase by 12.85307%.

For the LR overall test, the P-value=0.000<0.05. Thus, the null hypothesis that all estimated results are overall insignificant is rejected.

The individual Z-test suggest that x1 is insignificant at 95% confidence level as it has the p-values more than 0.05. However, x2,x3 and x4 are significant as its p-value are less than 0.05.

The counted R2 of the model is 0.735 > 0.05. From this R2 result, it can be said that the model fit the data quite well.

Perform the test to determine whether order probit is appropriated in this case? Give explanation why? (5 points).

```
. gologit2 y x1 x2 x3 x4, pl sto(oprobit) link(p)

Generalized Ordered Probit Estimates            Number of obs     =        170
                                                LR chi2(4)        =      81.28
                                                Prob > chi2       =     0.0000
Log likelihood = -107.19118                     Pseudo R2         =     0.2749

 ( 1)  [0]x1 - [1]x1 = 0
 ( 2)  [0]x2 - [1]x2 = 0
 ( 3)  [0]x3 - [1]x3 = 0
 ( 4)  [0]x4 - [1]x4 = 0
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |
          x1 |  -.1496765   .2384153    -0.63   0.530     -.616962    .3176089
          x2 |  -.5828778   .2454394    -2.37   0.018     -1.06393   -.1018254
          x3 |   .4722078   .2223359     2.12   0.034     .0364374    .9079783
          x4 |   .7043576   .0936777     7.52   0.000     .5207527    .8879624
       _cons |  -9.692317   1.418965    -6.83   0.000    -12.47344   -6.911197
-------------+----------------------------------------------------------------
1            |
          x1 |  -.1496765   .2384153    -0.63   0.530     -.616962    .3176089
          x2 |  -.5828778   .2454394    -2.37   0.018     -1.06393   -.1018254
          x3 |   .4722078   .2223359     2.12   0.034     .0364374    .9079783
          x4 |   .7043576   .0936777     7.52   0.000     .5207527    .8879624
       _cons |  -10.94181   1.474225    -7.42   0.000    -13.83124   -8.052381
------------------------------------------------------------------------------

. gologit2 y x1 x2 x3 x4, npl sto(goprobit) link(p)

Generalized Ordered Probit Estimates            Number of obs     =        170
                                                LR chi2(8)        =      86.46
                                                Prob > chi2       =     0.0000
Log likelihood = -104.60433                     Pseudo R2         =     0.2924
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
0						
x1	-.3791823	.3199502	-1.19	0.236	-1.006273	.2479086
x2	-.5173927	.3508267	-1.47	0.140	-1.205	.170215
x3	.1185091	.3271054	0.36	0.717	-.5226058	.7596239
x4	.5878705	.1212038	4.85	0.000	.3503154	.8254255
_cons	-7.612275	1.874432	-4.06	0.000	-11.28609	-3.938457
1						
x1	-.0379872	.2790866	-0.14	0.892	-.5849869	.5090125
x2	-.6366003	.2675227	-2.38	0.017	-1.160935	-.1122654
x3	.5847111	.2497659	2.34	0.019	.0951789	1.074243
x4	.745876	.1094295	6.82	0.000	.5313982	.9603538
_cons	-11.73092	1.740029	-6.74	0.000	-15.14131	-8.320524

. lrtest oprobit goprobit, stats

Likelihood-ratio test LR chi2(4) = 5.17
(Assumption: oprobit nested in goprobit) Prob > chi2 = 0.2699

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
oprobit	170	-147.8321	-107.1912	6	226.3824	245.1971
goprobit	170	-147.8321	-104.6043	10	229.2087	260.5667

Note: N=Obs used in calculating BIC; see [R] BIC note.

Order probit is appropriated in this case because we cannot reject the null hypothesis that is should be applied (P-value=0.2699>0.05).

From the data set “Final_q1-2_no.dta”:

In the study of decision to subscribe mobile phone services, three choices have been studied including AIS, DTAC, and TRUE.

(c) Estimate models for Y1i, Y2i, and Y3i assuming that the probability functions follow separate normal distribution function. Should we use these three separate probit models? Why? (4 points)

```
. use "C:\Users\user\Documents\BE TU\Year2\EE426_Final\Final_q1-2_9.dta"

. probit y1 x*, nolog

Probit regression                                 Number of obs     =        550
                                                  LR chi2(4)        =      34.33
                                                  Prob > chi2       =     0.0000
Log likelihood = -360.78596                       Pseudo R2         =     0.0454

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.433443    .405418     3.54   0.000     .6388383    2.228048
          x2 |  -.9249864   .3798049    -2.44   0.015     -1.66939   -.1805824
          x3 |   .6722486   .1793098     3.75   0.000     .3208078    1.023689
          x4 |    .305718   .1883133     1.62   0.104    -.0633692    .6748051
       _cons |  -.8274068   .3045617    -2.72   0.007    -1.424337   -.2304768
------------------------------------------------------------------------------

. probit y2 x*, nolog

Probit regression                                 Number of obs     =        550
                                                  LR chi2(4)        =      16.86
                                                  Prob > chi2       =     0.0021
Log likelihood = -360.47832                       Pseudo R2         =     0.0228

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.7999562   .3794418    -2.11   0.035    -1.543648    -.056264
          x2 |  -.5293842   .3648718    -1.45   0.147     -1.24452    .1857513
          x3 |  -.2681962   .1770488    -1.51   0.130    -.6152055    .0788132
          x4 |   .6654996   .1929992     3.45   0.001     .2872281    1.043771
```

```
       _cons |   .3625239   .3010094     1.20   0.228    -.2274436    .9524914
------------------------------------------------------------------------------


. probit y3 x*, nolog

Probit regression                                 Number of obs     =        550
                                                  LR chi2(4)        =      14.51
                                                  Prob > chi2       =     0.0058
Log likelihood = -329.56103                       Pseudo R2         =     0.0215

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.4782122   .3910551    -1.22   0.221    -1.244666    .2882418
          x2 |   .9566919    .378437     2.53   0.011     .2149691    1.698415
          x3 |  -.3740026   .1987409    -1.88   0.060    -.7635276    .0155224
          x4 |  -.4037282   .1849424    -2.18   0.029    -.7662086   -.0412478
       _cons |  -.5799735   .3165928    -1.83   0.067    -1.200484    .0405371
------------------------------------------------------------------------------
```

Though all separate three probit models are overall significant (look at the LR overall test) and are individually significant at the estimated result coded in green, we cannot be sure that they should be used.

We cannot tell from looking at these separate estimated results whether their disturbance terms are correlated or not. The test employ below in (d) should be conducted to determine whether the separate ones or MV Probit should be employed.

(d) Estimate models for Y1i, Y2i, and Y3i assuming that the probability functions follow multivariate normal probability distribution function (MV Probit models). Determine whether MVProbit is appropriated. Why? How does the MV Probit models differ from three separate probit models? (6 points)

```
. mvprobit(y1 x*)(y2 x*)(y3 x*), nolog

Multivariate probit (MSL, # draws = 5)            Number of obs     =        550
                                                  Wald chi2(12)     =      48.80
Log likelihood = -931.10282                       Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
```

```
          x1 |   1.272189   .3955664     3.22   0.001     .4968932    2.047485
          x2 |  -.7956449    .373739    -2.13   0.033     -1.52816     -.06313
          x3 |    .650857   .1800678     3.61   0.000     .2979307    1.003783
          x4 |   .2823215   .1873591     1.51   0.132    -.0848956    .6495386
       _cons |  -.7515269   .3063353    -2.45   0.014    -1.351933   -.1511208
-------------+----------------------------------------------------------------
y2           |
          x1 |  -.8718112   .3756413    -2.32   0.020    -1.608055   -.1355677
          x2 |  -.4706207   .3624317    -1.30   0.194    -1.180974    .2397323
          x3 |  -.2880062   .1777482    -1.62   0.105    -.6363862    .0603739
          x4 |   .6480574   .1924968     3.37   0.001     .2707705    1.025344
       _cons |   .4098823    .299649     1.37   0.171     -.177419    .9971837
-------------+----------------------------------------------------------------
y3           |
          x1 |  -.4187581   .3832823    -1.09   0.275    -1.169978    .3324614
          x2 |   .7598876   .3740493     2.03   0.042     .0267645    1.493011
          x3 |  -.2776699   .1866624    -1.49   0.137    -.6435215    .0881818
          x4 |  -.3461566   .1887768    -1.83   0.067    -.7161522    .0238391
       _cons |  -.5173486   .3177409    -1.63   0.103    -1.140109    .1054121
-------------+----------------------------------------------------------------
    /atrho21 |  -.5678597   .0739008    -7.68   0.000    -.7127025   -.4230168
-------------+----------------------------------------------------------------
    /atrho31 |  -.4442256    .075367    -5.89   0.000    -.5919422   -.2965091
-------------+----------------------------------------------------------------
    /atrho32 |  -.3381056   .0708779    -4.77   0.000    -.4770237   -.1991875
-------------+----------------------------------------------------------------
       rho21 |  -.5137857   .0543928    -9.45   0.000    -.6123687   -.3994689
-------------+----------------------------------------------------------------
       rho31 |  -.4171409   .0622526    -6.70   0.000     -.531291   -.2881147
-------------+----------------------------------------------------------------
       rho32 |  -.3257851   .0633552    -5.14   0.000    -.4438568   -.1965943
------------------------------------------------------------------------------
Likelihood ratio test of  rho21 = rho31 = rho32 = 0:

             chi2(3) =  239.445   Prob > chi2 = 0.0000
```

Since the p-value of the LR test of rho is less than 0.05, the hypothesis that there is no correlation among the disturbance terms of these three models can be rejected. Thus, MVProbit is appropriated in this case.

MV Probit differs from the three separate models such that there exists correlation among the disturbances term (the covariance shown in the variance-covariance matrix does not equal to zero). The decision on one choice affects another.

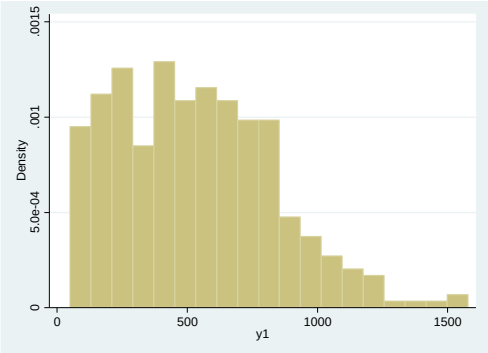
In the separate models, the correlation between the disturbance terms equals to zero. The decision on one choice is independent of another.

Looking at the nature of our study (decision to subscribe mobile phone service), it even makes sense to use MVProbit as the decision to buy AIS or not would be dependable on the decision to buy TRUE or DTAC. They are alternative choices and users usually choose only one.

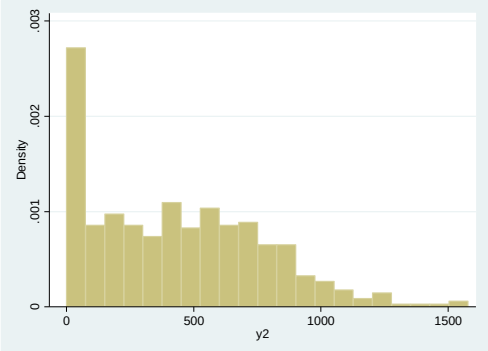
2) 2. From the data set “Final_q2_no.dta”:

(a) Plot histogram of y1i, y2i, y3i, compute descriptive statistics of these three variables. Determine limitations of these three dependent variables. (5 points)

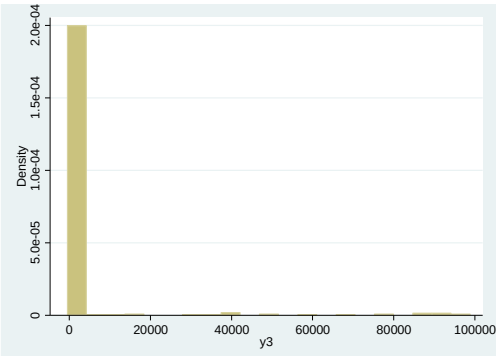
Histogram of y1i



Histogram of y2i

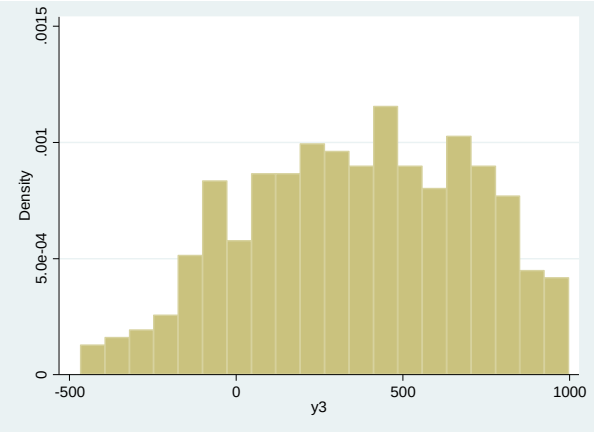


Histogram of y3i



. histogram y3 if y3<5000

(bin=20, start=-466.00415, width=73.209753)



. tabstat x y1 y2 y3, stat(N mean med sd min max)

stats	x	y1	y2	y3
N	450	366	450	450
mean	3.031962	530.4872	432.4673	3426.143
p50	2.996277	505.0885	409.4586	409.4586
sd	1.046268	301.9666	340.7485	14633.64
min	.2858976	50.21759	0	-466.0042
max	6.098752	1578.51	1578.51	98951.63

. sum y1 y2 y3

Variable	Obs	Mean	Std. Dev.	Min	Max
y1	366	530.4872	301.9666	50.21759	1578.51
y2	450	432.4673	340.7485	0	1578.51
y3	450	3426.143	14633.64	-466.0042	98951.63

From the histogram and the descriptive stats, y1 may have truncated problem at around 50, y2 may have censored problem at 0 and y3 may have outlier problem.

(b) Estimate the model (3) for y1i using OLS and truncated regression model. Determine the most appropriated model in this case. Why? What is the major problem in this case? What will happen if we ignore the problem? (5 points)

. *Estimate y1i

```
. *Using OLS
. reg y1 x

      Source |       SS           df       MS      Number of obs   =       366
-------------+----------------------------------   F(1, 364)       =    134.84
       Model |   8996193.7         1   8996193.7   Prob > F        =    0.0000
    Residual |  24285895.8       364   66719.494   R-squared       =    0.2703
-------------+----------------------------------   Adj R-squared   =    0.2683
       Total |  33282089.5       365  91183.8069   Root MSE        =     258.3

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   165.1108   14.21911    11.61   0.000     137.1489    193.0728
       _cons |   -5.70249   48.10933    -0.12   0.906    -100.3096    88.90464
------------------------------------------------------------------------------

. est store m_y1
. sum y1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y1 |        366    530.4872    301.9666   50.21759    1578.51

. scalar miny1=round(r(min))

*Estimate using the truncated model
. truncreg y1 x, ll(miny1) nolog
(note: 0 obs. truncated)

Truncated regression
Limit:   lower =         50                     Number of obs     =        366
         upper =       +inf                     Wald chi2(1)      =     103.78
Log likelihood = -2518.2088                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
```

```
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   228.2425   22.40469    10.19   0.000     184.3301    272.1548
       _cons |  -285.9187   85.48409    -3.34   0.001    -453.4644    -118.373
-------------+----------------------------------------------------------------
      /sigma |   305.1047   16.66777    18.31   0.000     272.4365     337.773
------------------------------------------------------------------------------
```

```
. est store y_y1_trunc
```

```
. *Test whether we should employ the truncated model
```

```
. lrtest m_y1 y_y1_trunc, force

Likelihood-ratio test                                 LR chi2(1)  =     65.86
(Assumption: m_y1 nested in y_y1_trunc)               Prob > chi2 =    0.0000
```

According to our estimated result, the truncated model should be used because we can reject the null hypothesis of the significant LR test that we should not employ the truncated model (p-value=0.000<0.05).

The truncated model should be employed in this case because there are missing observations – some Xs and Ys cannot be observed, but they exist.

This problem could stem from 1) survey design – we drop some observation out as they are not good representatives of the whole population and 2) the sample decides not to be in the sample themselves.

For the second case, the example is that we may not be able to collect information of all observations related to our study of nurses' salaries because some of our observations (=those who can work as nurses) may be offered a lower salary than expected, which result in them not choosing to be nurses.

The main problem is that the sample selection is selected based only on the value of y, causing a sample selection biased.

If we ignore this problem and continue to estimate using OLS method, our estimated results would be biased, stemming from the sample selection biased.

(c) Estimate the model (3) for y2i using OLS and Tobit model. Determine the most appropriated model in this case. Why? What is the major problem in this case? What will happen if we ignore the problem? (5 points)

```
. *Estimate y2i
```

```
. *Using OLS
```

```
. reg y2 x

      Source |       SS           df       MS      Number of obs   =       450
-------------+----------------------------------   F(1, 448)       =    252.45
       Model |  18789608.8         1  18789608.8   Prob > F        =    0.0000
```

```
    Residual |  33343584.7       448  74427.6445   R-squared       =    0.3604
-------------+----------------------------------   Adj R-squared   =    0.3590
       Total |  52133193.5       449  116109.562   Root MSE        =    272.81

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   195.5207   12.30555    15.89   0.000     171.3369    219.7044
       _cons |  -160.3439   39.46425    -4.06   0.000    -237.9019   -82.78586
------------------------------------------------------------------------------

. est store m_y2

. *Tobit
. tobit y2 x, ll(0)

Tobit regression                                  Number of obs     =        450
                                                  LR chi2(1)        =     212.41
                                                  Prob > chi2       =     0.0000
Log likelihood = -2788.9297                       Pseudo R2         =     0.0367

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   233.6658   14.78517    15.80   0.000     204.6091    262.7226
       _cons |  -307.4588   48.27432    -6.37   0.000    -402.3305   -212.5872
-------------+----------------------------------------------------------------
      /sigma |    306.134   11.36037                      283.808    328.4601
------------------------------------------------------------------------------
            68  left-censored observations at y2 <= 0
           382     uncensored observations
             0 right-censored observations

. est store m_y2c

. *Test whether we should employ the tobit model
. lrtest m_y2 m_y2c, force
```

```
Likelihood-ratio test                                 LR chi2(1)  =    745.09
(Assumption: m_y2 nested in m_y2c)                    Prob > chi2 =    0.0000
```

According to our estimated result, the tobit model should be used because we can reject the null hypothesis of the significant LR test that we should not employ the tobit model (p-value=0.000<0.05).

The tobit model should be employed in this case because there are censored observations (some Ys are censored at 0).

The main problem is that the sample is censored based only on the value of y, causing a sample selection biased.

If we ignore this problem and continue to estimate using OLS method, our estimated results would be biased, stemming from the sample selection biased.

(d) Estimate the model (3) for y3i using OLS and Tobit model. Determine the most appropriated model in this case. Why? What is the major problem in this case? What will happen if we ignore the problem? (5 points)

```
. *Estimate y3i
. *Using OLS
. reg y3 x

      Source |       SS           df       MS      Number of obs   =       450
-------------+----------------------------------   F(1, 448)       =     34.94
       Model |  6.9564e+09         1  6.9564e+09   Prob > F        =    0.0000
    Residual |  8.9194e+10       448   199093880   R-squared       =    0.0723
-------------+----------------------------------   Adj R-squared   =    0.0703
       Total |  9.6150e+10       449   214143484   Root MSE        =     14110

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   3762.054   636.4477     5.91   0.000      2511.26    5012.847
       _cons |   -7980.26   2041.107    -3.91   0.000    -11991.59   -3968.928
------------------------------------------------------------------------------

. est store m_y3
. tobit y3 x, ul(5000)

Tobit regression                                  Number of obs     =       450
```

```
                                                LR chi2(1)        =     100.78
                                                Prob > chi2       =     0.0000
Log likelihood = -3601.6045                     Pseudo R2         =     0.0138

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   492.4205   46.45093    10.60   0.000     401.1323    583.7087
       _cons |  -867.8368   148.7129    -5.84   0.000    -1160.096    -575.577
-------------+----------------------------------------------------------------
      /sigma |   1025.426   35.95979                      954.7552    1096.096
------------------------------------------------------------------------------

              0  left-censored observations
            426     uncensored observations
             24 right-censored observations at y3 >= 5000
```

```
. est store m_y3o_t
```

```
*Test whether we should employ the tobit model
. lrtest m_y3 m_y3o_t, force

Likelihood-ratio test                                 LR chi2(1)  =   2671.01
(Assumption: m_y3 nested in m_y3o_t)                  Prob > chi2 =    0.0000
```

According to our estimated result, the tobit model should be used because we can reject the null hypothesis of the significant LR test that we should not employ the tobit model (p-value=0.000<0.05).

The tobit model should be employed in this case because there are outlier observations (this is the main problem). The prediction would not be a good representative of what we want to study. Thus, we should censor the outlier observations at 5000 and employ the tobit model.

When we censor the data, the problem would be that the sample is censored based only on the value of y, causing a sample selection biased.

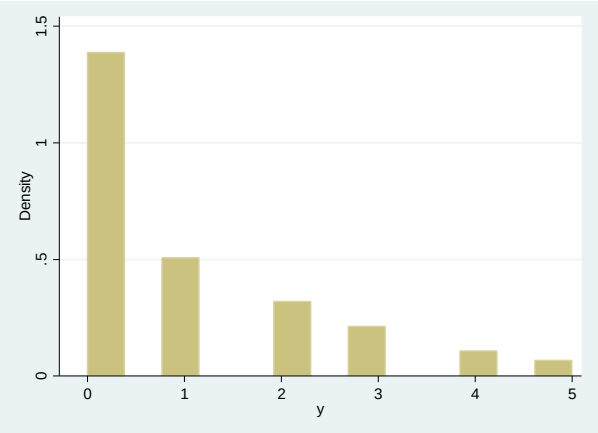
If we ignore this problem and continue to estimate using OLS method, our estimated results would be biased, stemming from the sample selection biased.

3) From the data set “Final_q3_no.dta”:

```
. reg y x1 x2 x3 x4

      Source |       SS           df       MS      Number of obs   =       195
```

```
-------------+----------------------------------   F(4, 190)       =      5.20
       Model |  33.9287467         4  8.48218668   Prob > F        =    0.0005
    Residual |  309.989202       190  1.63152212   R-squared       =    0.0987
-------------+----------------------------------   Adj R-squared   =    0.0797
       Total |  343.917949       194  1.77277293   Root MSE        =    1.2773

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0736619   .0447717     1.65   0.102    -.0146516    .1619754
          x2 |   .1320037    .046928     2.81   0.005     .0394369    .2245706
          x3 |   .2009962   .0671201     2.99   0.003        .0686    .3333925
          x4 |  -.0332937   .0482185    -0.69   0.491    -.1284059    .0618186
       _cons |   .9218831   .1082691     8.51   0.000     .7083192    1.135447
------------------------------------------------------------------------------
```

```
. histogram y
(bin=13, start=0, width=.38461538)
```



According to the histogram, there is limitation of dependent variable as the dependent variables are counted number (non-negative integer), which may be Poisson, negative binomial, zero inflated Poisson, or zero inflated negative binomial. We should conduct further test to see exactly what it is.

(b) Estimate models for Yi assuming that the probability functions follow Poisson probability distribution. Perform GOF test and determine whether Poisson is appropriated in this case. Interpret the estimated result (sign and meaning (in term of incidence-rate ratios), overall test, individual test, pseudo R2, marginal effects). (5 points)

```
. poisson y x1 x2 x3 x4, nolog
```

```
Poisson regression                              Number of obs     =        195
                                                LR chi2(4)        =      35.13
                                                Prob > chi2       =     0.0000
Log likelihood = -272.05844                     Pseudo R2         =     0.0607

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0747825   .0346503     2.16   0.031     .0068692    .1426957
          x2 |   .1345546   .0365751     3.68   0.000     .0628688    .2062405
          x3 |   .2161646   .0550499     3.93   0.000     .1082687    .3240605
          x4 |  -.0344743    .038076    -0.91   0.365    -.1091018    .0401533
       _cons |   -.170492   .0938452    -1.82   0.069    -.3544253    .0134413
------------------------------------------------------------------------------
```

```
. estat gof

         Deviance goodness-of-fit =  313.9894
         Prob > chi2(190)         =    0.0000

         Pearson goodness-of-fit  =  331.8099
         Prob > chi2(190)         =    0.0000
```

According the the GOF test, the null hypothesis that Poisson should be applied is rejected. Thus, the Poisson regression is inappropriate.

```
. poisson y x1 x2 x3 x4, ir nolog

Poisson regression                              Number of obs     =        195
                                                LR chi2(4)        =      35.13
                                                Prob > chi2       =     0.0000
Log likelihood = -272.05844                     Pseudo R2         =     0.0607

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |    1.07765   .0373408     2.16   0.031     1.006893    1.153379
```

```
          x2 |   1.144027   .0418429     3.68   0.000     1.064887    1.229049
          x3 |   1.241307   .0683338     3.93   0.000     1.114347    1.382731
          x4 |   .9661132   .0367857    -0.91   0.365     .8966391     1.04097
       _cons |   .8432498    .079135    -1.82   0.069     .7015765    1.013532
------------------------------------------------------------------------------


. est store poisson


. mfx


Marginal effects after poisson
      y  = Predicted number of events (predict)
         =  .89175299
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0666875      .03059    2.18   0.029   .006741  .126634  -.278437
      x2 |   .1199895      .03168    3.79   0.000   .057905  .182074   .838739
      x3 |   .1927654      .04717    4.09   0.000   .100319  .285212  -.292205
      x4 |  -.0307425      .03388   -0.91   0.364  -.097155   .03567  -.784821
------------------------------------------------------------------------------
```

Interpreting the result from the Poisson regression:

X1, x2 and x3 have a positive sign while x4 has a negative sign. We should check these signs with the theory.

IRR tells us that…

If x1 changes by 1 unit, y will change 1.07765 units (7.765%).

If x2 changes by 1 unit, y will change 1.144027 units (14.4027%).

If x3 changes by 1 unit, y will change 0.9661132 unit.

If x4 changes by 1 unit, y will change 0.0367857 unit.

Noted that if the IRR>1, the change is positive. If the IRR<1, the change is negative.

From the LR test above as the p-value is 0.0000 < 0.05, we can reject the null hypothesis that the estimated results are overall insignificant. All the estimated results are overall significant.

The individual Z-test suggest that for x1 and x4 are insignificant at 95% confidence level as they presented the p-values more than 0.05. x2 and x3 are significant as its p-value is less than 0.05.

The Pseudo R2 of the estimated result is 0.0607. This R2 cannot be used to make interpretation. It can only be used for comparison. However, the number is quite low, we may be able to state roughly that the estimated results do not quite fit the data.

In addition, the marginal effects at mean tells us that x1, x2 and, x3 has a positive mfx sign(IRR>1) and x4 has a negative mfx sign(IRR<1). This is correct as it goes along with the IRR sign.

(c) Estimate models for Yi assuming that the probability functions follow Negative Binomial probability distribution. Determine whether Negative Binomial regression model is appropriated in this case. Interpret your estimated result (sign and meaning (in term of incidence-rate ratios), overall test, individual test, pseudo R 2, marginal effects). (5 points)

```
. nbreg y x1 x2 x3 x4, nolog

Negative binomial regression                    Number of obs     =        195
                                                LR chi2(4)        =      18.78
Dispersion     = mean                           Prob > chi2       =     0.0009
Log likelihood = -258.03883                     Pseudo R2         =     0.0351

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0991694     .05265     1.88   0.060    -.0040226    .2023615
          x2 |    .149569   .0518563     2.88   0.004     .0479326    .2512055
          x3 |   .2036212   .0700912     2.91   0.004      .066245    .3409974
          x4 |  -.0389973   .0512817    -0.76   0.447    -.1395076    .0615131
       _cons |  -.1881754   .1228355    -1.53   0.126    -.4289285    .0525778
-------------+----------------------------------------------------------------
    /lnalpha |   -.198206   .2968966                     -.7801126    .3837006
-------------+----------------------------------------------------------------
       alpha |   .8202009   .2435148                      .4583544    1.467706
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =   28.04 Prob>=chibar2 = 0.000

. nbreg y x1 x2 x3 x4, ir nolog

Negative binomial regression                    Number of obs     =        195
                                                LR chi2(4)        =      18.78
Dispersion     = mean                           Prob > chi2       =     0.0009
Log likelihood = -258.03883                     Pseudo R2         =     0.0351

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
          x1 |   1.104253   .0581389     1.88   0.060     .9959855     1.22429
          x2 |   1.161334   .0602224     2.88   0.004       1.0491    1.285574
          x3 |   1.225834   .0859201     2.91   0.004     1.068489     1.40635
          x4 |   .9617533   .0493204    -0.76   0.447     .8697864    1.063444
       _cons |   .8284694   .1017655    -1.53   0.126     .6512065    1.053985
-------------+----------------------------------------------------------------
    /lnalpha |   -.198206   .2968966                     -.7801126    .3837006
-------------+----------------------------------------------------------------
       alpha |   .8202009   .2435148                      .4583544    1.467706
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =   28.04 Prob>=chibar2 = 0.000
```

According to the LR test of alpha, we can reject the null hypothesis of the test as 0.000<0.05. Rejection means that negative binomial is appropriated.

```
. est store nb
```

```
. mfx
```

```
Marginal effects after nbreg
      y  = Predicted number of events (predict)
         =  .88760226
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |    .088023      .04662    1.89   0.059  -.003346  .179392  -.278437
      x2 |   .1327578      .04583    2.90   0.004   .042928  .222587   .838739
      x3 |   .1807347       .0619    2.92   0.004   .059422  .302048  -.292205
      x4 |  -.0346141      .04551   -0.76   0.447   -.12381  .054582  -.784821
------------------------------------------------------------------------------
```

Interpreting the result from the NB regression:

X1, x2 and x3 have a positive sign while x4 has a negative sign. We should check these signs with the theory.

IRR tells us that…

If x1 changes by 1 unit, y will change 1.104253 units.

If x2 changes by 1 unit, y will change 1.161334 units.

If x3 changes by 1 unit, y will change 1.225834 units.

If x4 changes by 1 unit, y will change 0.9617533 unit.

Noted that if the IRR>1, the change is positive. If the IRR<1, the change is negative.

From the LR test above as the p-value is 0.0009 < 0.05, we can reject the null hypothesis that the estimated results are overall insignificant. All the estimated results are overall significant.

The individual Z-test suggest that only x4 is insignificant at 95% confidence level as it presented the p-values more than 0.05. Others are significant.

The Pseudo R2 of the estimated result is 0.0351. This R2 cannot be used to make interpretation. It can only be used for comparison. However, the number is quite low, we may be able to state roughly that the estimated results do not quite fit the data.

In addition, the marginal effects at mean tells us that x1, x2 and, x3 has a positive mfx sign(IRR>1) and x4 has a negative mfx sign(IRR<1). This is correct as it goes along with the IRR sign.

(d) Estimate models for Yi assuming that the model is Zero Inflated Poisson (x1i, x2i, and x3i are independent variables in Poisson model and x4i is independent variable in Inflated (Logit) model). Interpret your estimated result. Determine which model (Linear regression model, Poisson, Negative Binomial, or ZIP) is the most appropriated model in this case? Why? (provide the tests) (7 points)

```
. zip y x1 x2 x3, inflate(x4) vuong nolog

Zero-inflated Poisson regression                Number of obs     =        195
                                                Nonzero obs       =         91
                                                Zero obs          =        104

Inflation model = logit                         LR chi2(3)        =      13.97
Log likelihood  = -257.1601                     Prob > chi2       =     0.0029

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
          x1 |   .0838201   .0439048     1.91   0.056    -.0022318    .1698719
          x2 |    .118921   .0407701     2.92   0.004     .0390131    .1988288
          x3 |   .1510494   .0562254     2.69   0.007     .0408497    .2612491
       _cons |   .3107383   .1181343     2.63   0.009     .0791992    .5422774
-------------+----------------------------------------------------------------
inflate      |
          x4 |   .0341237   .1044039     0.33   0.744    -.1705041    .2387515
       _cons |  -.5377385   .2668255    -2.02   0.044    -1.060707   -.0147701
------------------------------------------------------------------------------
Vuong test of zip vs. standard Poisson:            z =     2.41  Pr>z = 0.0080
```

According to the Vuong test, we can reject the null hypothesis that Poisson is more appropriate as 0.0080<0.5. Thus, ZIP is a more appropriated model in this case.

```
. mfx
```

```
Marginal effects after zip
      y  = Predicted number of events (predict)
         =  .89834718
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0752995      .03919    1.92   0.055  -.001516  .152115  -.278437
      x2 |   .1068323      .03592    2.97   0.003    .03643  .177235   .838739
      x3 |   .1356948      .04935    2.75   0.006   .038968  .232422  -.292205
      x4 |  -.0111125      .03416   -0.33   0.745  -.078067  .055842  -.784821
------------------------------------------------------------------------------
```

Interpreting the result from the ZIP regression:

X1, x2, x3 and x4 have a positive sign. We should check these signs with the theory.

From the LR test above as the p-value is 0.0029 < 0.05, we can reject the null hypothesis that the estimated results are overall insignificant. All the estimated results are overall significant.

The individual Z-test suggest that only x4 is insignificant at 95% confidence level as it presented the p-values more than 0.05. Others are significant.

In addition, the marginal effects at mean tells us that x1, x2 and, x3 has a positive mfx (x moves in a positive direction with y) and x4 has a negative mfx (x moves in a negative direction with y).

Though the LR test of alpha suggest that the Negative Binomial is more appropriated than the Poisson.

The most appropriated model in this case is the ZIP model as suggested in the Vuong test and the histogram (having excess zeros).

4. From the data set “Final_q4_no.dta”:

(a) Test whether the series yt and xt are stationary series and determine their order of integration. Give explanation of the process of the test. Why is testing equation important? Why is stationary property important to time series analysis? (5 points)

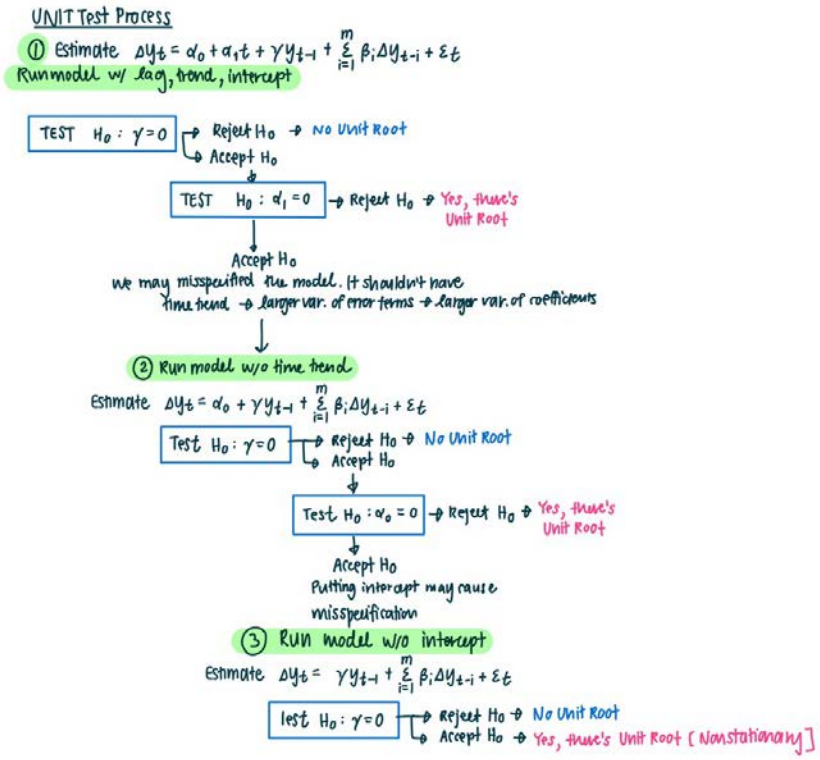
```
. tsset t
        time variable:  t, 1 to 500
                delta:  1 unit


. *Test series yt


. *Test with intercept, trend, lags


. dfuller y, trend lags(1) regress
```

```
Augmented Dickey-Fuller test for unit root         Number of obs   =       498

                                 ---------- Interpolated Dickey-Fuller ---------
                 Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -15.063            -3.980            -3.420            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
D.y          |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           y |
         L1. |  -.9767421   .0648458   -15.06   0.000     -1.10415   -.8493345
         LD. |  -.0595469   .0449891    -1.32   0.186    -.1479405    .0288466
      _trend |   .0017552   .0021074     0.83   0.405    -.0023853    .0058957
       _cons |   .8014772   .6104414     1.31   0.190    -.3979045    2.000859
------------------------------------------------------------------------------

. *Test series xt

. *Test with intercept, trend, lags

. dfuller x, trend lags(1) regress

Augmented Dickey-Fuller test for unit root         Number of obs   =       498

                                 ---------- Interpolated Dickey-Fuller ---------
                 Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -15.216            -3.980            -3.420            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
```

```
D.x          |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |
         L1. |  -.9880949   .0649366   -15.22   0.000    -1.115681   -.8605089
         LD. |  -.0503038   .0449959    -1.12   0.264    -.1387108    .0381032
      _trend |   .0027802   .0030099     0.92   0.356    -.0031336     .008694
       _cons |   .3732585   .8684657     0.43   0.668    -1.333084    2.079601
------------------------------------------------------------------------------
```

Both yt and xt are stationary series according to the unit root test. We can reject the hull hypothesis that there is a unit root, which is the sign of non-stationary series. Their order of integration is both I(0).

Testing for stationary has the process as follows…



Stationary is important to time series analysis because stationarity means that the statistical properties of a time series (or rather the process generating it) do not change over time.

Many useful analytical tools and statistical tests and models rely on it. It would be hard to estimate things that change its properties all the time.

(b) Estimate Autoregressive Integrated Moving Average (ARIMA(p,d,q)) model for yt – determine the most appropriated order for p, d, and q using SBIC given the maximum lag equals 4. Make dynamic forecast for period time = 501 to 505. (5 points)

```
. do "C:\Users\user\AppData\Local\Temp\STD00000000.tmp"
```

```
. qui arima y, arima(1,0,1) nolog

. est store arima101

. qui arima y, arima(1,0,2) nolog

. est store arima102

. qui arima y, arima(1,0,3) nolog

. est store arima103

. qui arima y, arima(1,0,4) nolog

. est store arima104

.
. qui arima y, arima(2,0,1) nolog

. est store arima201

. qui arima y, arima(2,0,2) nolog

. est store arima202

. qui arima y, arima(2,0,3) nolog

. est store arima203

. qui arima y, arima(2,0,4) nolog

. est store arima204

.
. qui arima y, arima(3,0,1) nolog

. est store arima301
```

```
. qui arima y, arima(3,0,2) nolog

. est store arima302

. qui arima y, arima(3,0,3) nolog

. est store arima303

. qui arima y, arima(3,0,4) nolog

. est store arima304

.
. qui arima y, arima(4,0,1) nolog

. est store arima401

. qui arima y, arima(4,0,2) nolog

. est store arima402

. qui arima y, arima(4,0,3) nolog

. est store arima403

. qui arima y, arima(4,0,4) nolog

. est store arima404

.
. est table arima10*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

-------------------------------------------------------------------------------
    Variable |   arima101        arima102        arima103        arima104
-------------+-----------------------------------------------------------------
y            |
       _cons |  1.2746367***    1.2773443***     1.275404***    1.2758015***
-------------+-----------------------------------------------------------------
```

```
ARMA          |
          ar  |
          L1. | -.84980825***    .82099212***   -.85104923***   -.80807326***
              |
          ma  |
          L1. |  .80521742***     -.860762***    .81604959***    .77324155***
          L2. |                  .07394483       .02682787        .0263373
          L3. |                                  .01836334       .04923956
          L4. |                                                  .04840352
--------------+----------------------------------------------------------------
sigma         |
       _cons  |  6.7248768***    6.7248695***    6.7231267***    6.7158723***
--------------+----------------------------------------------------------------
Statistics    |
           N  |        500             500             500             500
          ll  | -1662.3879       -1662.387      -1662.2564      -1661.7195
        chi2  |  49.994347       24.262953       50.181672       41.322777
         aic  |  3332.7759       3334.7739       3336.5129        3337.439
         bic  |  3349.6343       3355.8469       3361.8005       3366.9413
------------------------------------------------------------------------------
                                             legend: * p<.1; ** p<.05; *** p<.01

. est table arima20*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

------------------------------------------------------------------------------
    Variable |   arima201        arima202        arima203        arima204
-------------+----------------------------------------------------------------
y            |
       _cons |  1.2771689***    1.2764632***    1.2767652***    1.2767389***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
          L1. |  .73048354***    .03791482      -.00480599       -.0154302
          L2. |  .07850978       .72192023***    .72310514***    .68244375**
              |
          ma |
          L1. | -.77256153***   -.06202533      -.03305636      -.02147538
          L2. |                  -.6522344**    -.65538207**    -.62900465**
```

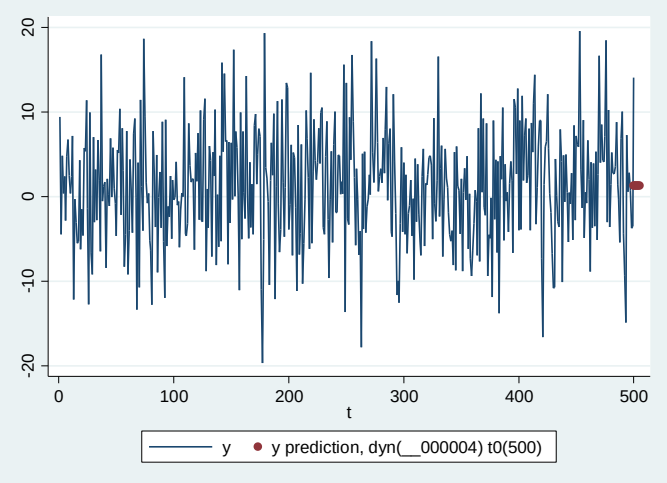
```
         L3. |                                  .02497671       .02348086
         L4. |                                                  .02543326
-------------+------------------------------------------------------------
sigma        |
       _cons |  6.7241162***    6.7134733***    6.7119091***    6.7102127***
-------------+------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.3251      -1661.5377      -1661.4163      -1661.2914
        chi2 |  20.849523       20.595779        21.73075       20.352835
         aic |  3334.6502       3335.0755       3336.8326       3338.5828
         bic |  3355.7232       3360.3631       3366.3349       3372.2997
------------------------------------------------------------------------------
                                             legend: * p<.1; ** p<.05; *** p<.01

. est table arima30*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

------------------------------------------------------------------------------
    Variable |   arima301        arima302        arima303        arima304
-------------+----------------------------------------------------------------
y            |
       _cons |  1.2755297***    1.2769634***    1.2933606***     1.293271***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.84284581**    -.03682361       1.0744837***    1.0730719***
         L2. |    .030008       .72458986***    .63729617**      .6401829
         L3. |  .02039265        .0255585      -.82706222***   -.82861588***
             |
          ma |
         L1. |  .80941431**     .00001219      -1.1343109      -1.1340837
         L2. |                 -.65768012**    -.53586248      -.53806189
         L3. |                                  .79575601       .79934309
         L4. |                                                 -.00173603
-------------+----------------------------------------------------------------
sigma        |
       _cons |  6.7230454***    6.7118075***    6.6269444       6.6269383
-------------+----------------------------------------------------------------
```

```
Statistics   |
           N |        500             500             500             500
          ll | -1662.2439      -1661.4251      -1657.1392      -1657.1385
        chi2 |  51.593025       22.250831       34788.218       34957.834
         aic |  3336.4877       3336.8502       3328.2783        3330.277
         bic |  3361.7754       3366.3524       3357.7806       3363.9939
-------------------------------------------------------------------------------
                                              legend: * p<.1; ** p<.05; *** p<.01

. est table arima40*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

-------------------------------------------------------------------------------
    Variable |   arima401        arima402        arima403        arima404
-------------+-----------------------------------------------------------------
y            |
       _cons |  1.2761116***    1.2770609***     1.293265***    1.2796961***
-------------+-----------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.76243576**    -.05075943       1.0709771***   -.05534888
         L2. |   .0310487       .64656215*       .6423814       -.2819904
         L3. |  .05134517       .02572557      -.82718262***    .08739219
         L4. |  .05038703       .02752599      -.00178093       .71157237***
             |
          ma |
         L1. |  .72793952**     .01349866      -1.1319765       .02621095
         L2. |                 -.59376365      -.54039239       .35697385
         L3. |                                  .79809934      -.09739485
         L4. |                                                 -.63931632
-------------+-----------------------------------------------------------------
sigma        |
       _cons |  6.7154317***    6.7100257***    6.6269182       6.6604571
-------------+-----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll |  -1661.697      -1661.2936      -1657.1385       -1659.124
        chi2 |  34.357426       19.568757       48156.523       81127.242
         aic |   3337.394       3338.5872       3330.2771        3338.248
```

```
         bic |  3366.8962        3372.304       3363.9939       3380.3941
-----------------------------------------------------------------------------
                                              legend: * p<.1; ** p<.05; *** p<.01
```

The most appropriated order for p,d,q is 1,0,1 as it as the lowest SBIC.

```
*Dynamic Forecast for period time=501 to 505

. est restore arima101
(results arima101 are active now)

. set obs 505
number of observations (_N) was 500, now 505

. replace t=_n
(5 real changes made)

. predict yhat, y dynamic(.) t0(500)
Note: beginning dynamic predictions in period         3.
(499 missing values generated)

. twoway (line y t, sort)(scatter yhat t, sort)
```



(c) Estimate model (5) using OLS by employing yt as dependent variable and xt as explanatory variable, determine whether ARCH-effect significantly occurs, and state null hypothesis of the test. Why the test is classified as LM test? (3 points)

```
. *OLS

. reg y x
```

```
      Source |       SS           df       MS      Number of obs   =       500
-------------+----------------------------------   F(1, 498)       =  57334.19
       Model |  22575.9547         1  22575.9547   Prob > F        =    0.0000
    Residual |  196.092864       498  .393760772   R-squared       =    0.9914
-------------+----------------------------------   Adj R-squared   =    0.9914
       Total |  22772.0475       499  45.6353658   Root MSE        =     .6275

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .6975582   .0029132   239.45   0.000     .6918344    .7032819
       _cons |   .5171023   .0282414    18.31   0.000     .4616154    .5725892
------------------------------------------------------------------------------


. estat archlm

LM test for autoregressive conditional heteroskedasticity (ARCH)

-----------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+---------------------------------------------------------------
       1     |         38.575               1                   0.0000
-----------------------------------------------------------------------------

         H0: no ARCH effects      vs.  H1: ARCH(p) disturbance
```

There exist significant ARCH effects since p-value of the ARCHLM test is 0.000<0.05. We can reject H0. The null hypothesis is there is no ARCH effects.

This test is classified as LM test because it uses only the restricted model in the estimation process.

(d) Estimate GARCH(p,q) for yt using xt as explanatory variable for mean equation (model (5) and (6)) – determine the most appropriated order p and q for variance equation using SBIC given the maximum lag equals to 2. Then, predict the variance of yt using the estimated result of GARCH(p,q) model with the most appropriated lag. (4 points)

```
. do "C:\Users\user\AppData\Local\Temp\STD00000000.tmp"


. qui arch y x, arch(1/1) garch (1/1)


. est store garch11


. qui arch y x, arch(1/1) garch (1/2)
```

```
. est store garch12

. qui arch y x, arch(1/2) garch (1/1)

. est store garch21

. qui arch y x, arch(1/2) garch (1/2)

. est store garch22

. est table garch*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

-------------------------------------------------------------------------------
    Variable |    garch11         garch12         garch21         garch22
-------------+-----------------------------------------------------------------
y            |
           x |  .69773869***     .6977268***    .69772205***    .69760958***
       _cons |  .51847361***    .51845244***    .51843288***    .51601408***
-------------+-----------------------------------------------------------------
ARCH         |
        arch |
         L1. |  .36232686***     .3609696***    .36060785***    .37252464***
         L2. |                                  .01569003       .30968046**
             |
       garch |
         L1. |  .31565892***    .32680084*      .28713554       -.3957941
         L2. |                 -.00999513                        .1090091
             |
       _cons |  .12883641***     .1289152***    .13463808*      .24600813***
-------------+-----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -443.74734      -443.74184      -443.74008      -442.93095
        chi2 |  72729.328       72790.406       72823.901       74328.442
         aic |  897.49468       899.48368       899.48017       899.86189
         bic |  918.56772       924.77133       924.76782       929.36415
```

```
------------------------------------------------------------------------------
                                                legend: * p<.1; ** p<.05; *** p<.01
```

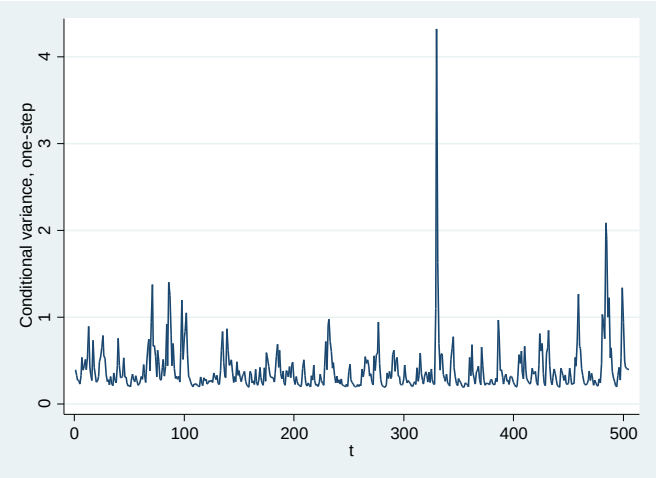
.

end of do-file

According to the estimation result above, garch (1,1) is the most appropriated order for the variance equation as it has the lowest SBIC.

***Predict the variance of yt using garch(1,1)**

```
. est restore garch11
(results garch11 are active now)

. predict sigma2, variance

. line sigma2 t
```



(e) Among these three models, ARCH(1), GARCH(1,1), or EGARCH(1,1,1), which model is the most appropriated in this case? Why? What are the differences among ARCH, GARCH, and EGARCH? (3 points)

I think the most appropriated model is GARCH(1,1) or EGARCH because the data seems to have a volatile variance. With high frequency TS data, things that happen yesterday may affect our decision today.

ARCH differs from GARCH and EGARCH such that it does not incorporate the conditional variance equation into the model.

GARCH differs from EGACRH such that GARCH did not take in account that reactions are not always happening at the same degree.

The degree of reactions varies according to the situation, such as good news and bad news. Bad news may affect people decision more. EGARCH takes this point into consideration and try to smoothen the reactions as much as possible.

5. From the data set "Final_q5_no.dta":

(a) Estimate VARs models using yt and xt as endogenous variables and determine the most appropriated lags models using SBIC. (5 points).

(b) Perform stability test and Granger exogeneity test. Determine whether assumptions of VARs are satisfied, explain your evaluation criteria. If the stability assumption is unsatisfied, what will happen? (5 points)

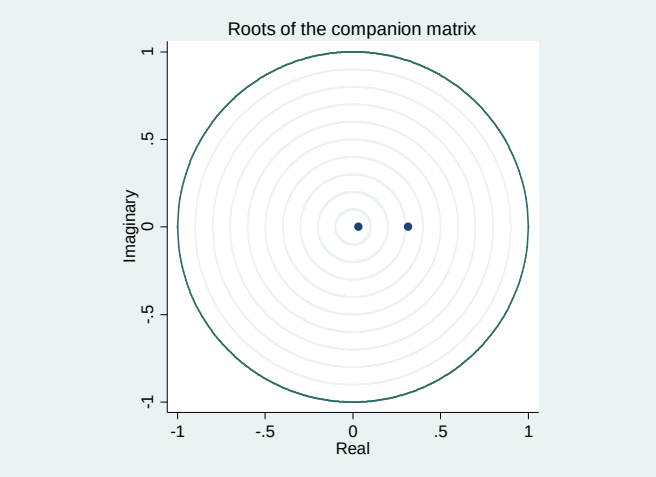
```
. use "C:\Users\user\Documents\BE TU\Year2\EE426_Final\Final_q5_9.dta", clear

. tsset t
        time variable:  t, 1 to 500
                delta:  1 unit

. varsoc y x

   Selection-order criteria
   Sample:  5 - 500                              Number of obs      =       496
  +---------------------------------------------------------------------------+
  |lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC    |
  |----+----------------------------------------------------------------------|
  |  0 |  -1366.8                      .855224   5.51936   5.52602   5.53632  |
  |  1 | -1319.84  93.915*   4  0.000  .719206*  5.34615*  5.36612*  5.39703* |
  |  2 | -1318.91  1.8705    4  0.760   .72815    5.3585    5.3918   5.44331  |
  |  3 | -1317.91  1.9948    4  0.737  .737021   5.37061   5.41722   5.48935  |
  |  4 | -1315.66  4.4998    4  0.343  .742244   5.37767   5.43759   5.53033  |
  +---------------------------------------------------------------------------+
   Endogenous:  y x
    Exogenous:  _cons

. *Stability Test

. var y x, lag(1/1)

Vector autoregression

Sample:  2 - 500                              Number of obs     =        499
Log likelihood =  -1329.952                   AIC               =   5.354515
FPE            =   .7252502                   HQIC              =   5.374393
```

```
Det(Sigma_ml)  =   .7080172                      SBIC              =   5.405168

Equation           Parms      RMSE     R-sq      chi2     P>chi2
------------------------------------------------------------------
y                      3      .99439   0.0147   7.447953   0.0241
x                      3     .997952   0.0887   48.54676   0.0000
------------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           y |
         L1. |   .1364689   .0500817     2.72   0.006     .0383105    .2346273
             |
           x |
         L1. |   .0536877   .0479556     1.12   0.263    -.0403036    .1476789
             |
       _cons |   .3380707   .0544817     6.21   0.000     .2312886    .4448528
-------------+----------------------------------------------------------------
x            |
           y |
         L1. |   .3428837   .0502611     6.82   0.000     .2443738    .4413937
             |
           x |
         L1. |   .2115588   .0481274     4.40   0.000     .1172308    .3058867
             |
       _cons |   .1233484   .0546768     2.26   0.024     .0161838     .230513
------------------------------------------------------------------------------

. varstable, graph

   Eigenvalue stability condition

  +--------------------------------------+
  |        Eigenvalue        |   Modulus   |
  |--------------------------+-------------|
  |   .3147911               |   .314791   |
```

```
|  .03323652               |   .033237   |
+----------------------------------------+
```

All the eigenvalues lie inside the unit circle.

VAR satisfies stability condition.



. *Granger Test

. vargranger

Granger causality Wald tests

```
+-------------------------------------------------------------------------+
|          Equation           Excluded |   chi2     df Prob > chi2 |
|-----------------------------------------+---------------------------|
|                 y                   x |  1.2533     1    0.263    |
|                 y                 ALL |  1.2533     1    0.263    |
|-----------------------------------------+---------------------------|
|                 x                   y |   46.54     1    0.000    |
|                 x                 ALL |   46.54     1    0.000    |
+-------------------------------------------------------------------------+
```

According to the stability test, the system is stable as all eigenvalues lie inside the unit circle.

Moreover, referring to the granger test. We can not reject the null hypothesis for the case of y causes x, but we can reject the null hypothesis for the case that x causes. Thus, this model does not guarantee endogeneity.

The VARs assumption of stability is satisfied, but the assumption of interdependence(endogeneity) does not. We usually have two criteria for VARs evaluation: stability and endogeneity.

If the stability is not satisfied, our prediction would not yield a good estimation results for the dynamic system.

(c) Perform Impulse response function analysis (irf), Orthogonal impulse response function analysis (oirf), Cumulative impulse response function analysis (coirf), make interpretation of the analysis, and determine which variable has more impact(using Cholesky order – yt xt). (5 points)

```
. irf create order1, o(y x) step(5) set(irf02)

(file irf02.irf now active)

(file irf02.irf updated)


. irf table irf, impulse(y x) response(y x)
```

Results from order1

	(1)	(1)	(1)	(2)	(2)	(2)
step	irf	Lower	Upper	irf	Lower	Upper
0	1	1	1	0	0	0
1	.136469	.03831	.234627	.342884	.244374	.441394
2	.037032	-.011434	.085499	.119333	.067879	.170787
3	.01146	-.008327	.031248	.037944	.013034	.062853
4	.003601	-.003888	.011091	.011957	.000534	.023379
5	.001133	-.001619	.003885	.003764	-.001085	.008614

	(3)	(3)	(3)	(4)	(4)	(4)
step	irf	Lower	Upper	irf	Lower	Upper
0	0	0	0	1	1	1
1	.053688	-.040304	.147679	.211559	.117231	.305887
2	.018685	-.014426	.051796	.063166	.025357	.100974
3	.005941	-.006277	.018159	.01977	.001807	.037733
4	.001872	-.002612	.006356	.00622	-.001539	.013978
5	.000589	-.001029	.002208	.001958	-.001169	.005084

95% lower and upper bounds reported

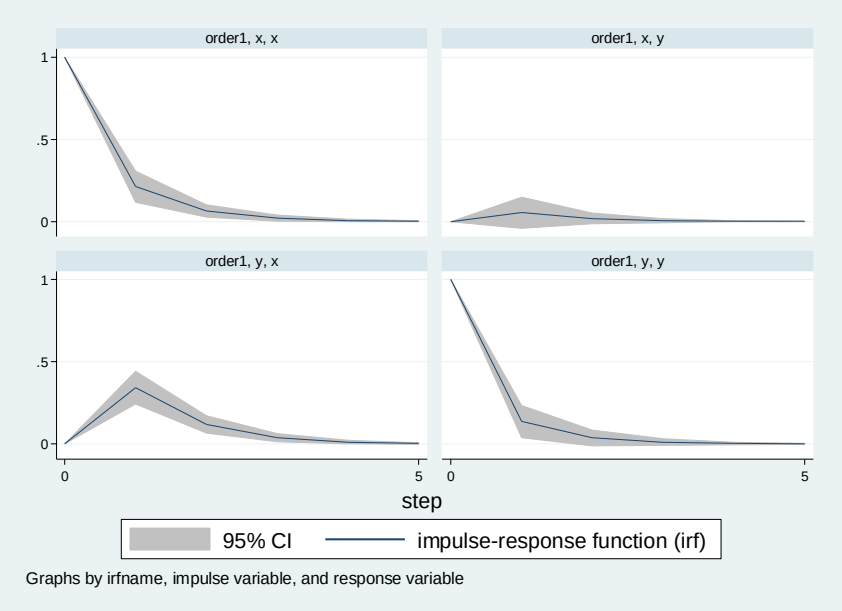
(1) irfname = order1, impulse = y, and response = y

(2) irfname = order1, impulse = y, and response = x

(3) irfname = order1, impulse = x, and response = y

(4) irfname = order1, impulse = x, and response = x

. irf graph irf, impulse(y x) response(y x)



Due to the graph above, all cases present a positive impact. The magnitude is uncertain because the size of the grey area is quite difficult to determine just by looking. The largest grey area represents the largest magnitude.

Only from the picture, all cases seem to have a stable system. The effect seems to die down when time passes.

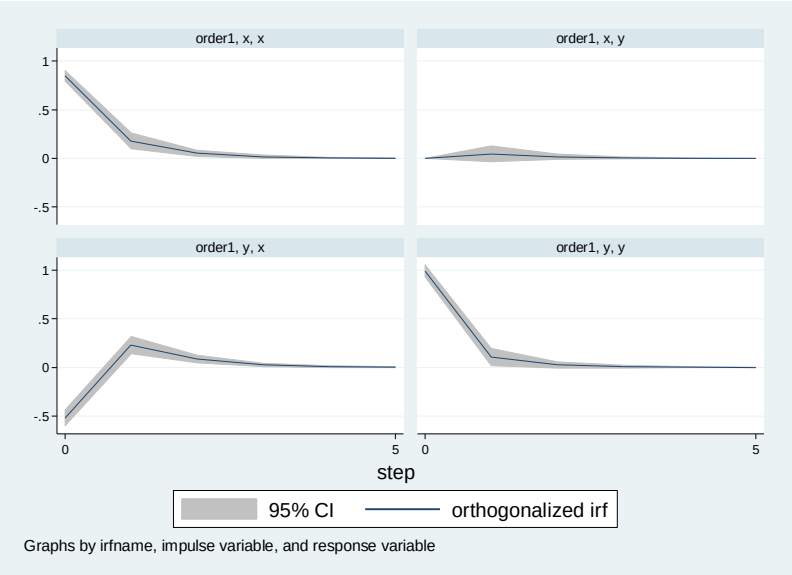
Unfortunately, some part of the grey area may cross the zero-horizontal line. This shows that they may be insignificant.

. irf table oirf, impulse(y x) response(y x)

Results from order1

	(1)	(1)	(1)	(2)	(2)	(2)
step	oirf	Lower	Upper	oirf	Lower	Upper
0	.991397	.929889	1.0529	-.519194	-.60033	-.438057
1	.107421	.020571	.19427	.230094	.140717	.319471
2	.027013	-.005379	.059405	.085511	.046266	.124756
3	.008277	-.005221	.021776	.027353	.009947	.044759

```
|4       | .002598     -.002552    .007748   | .008625     .00091      .016339   |
|5       | .000818     -.001084    .002719   | .002716     -.000561    .005992   |
+-------------------------------------------------------------------------------+


+-------------------------------------------------------------------------------+
|        |    (3)         (3)         (3)    |    (4)         (4)         (4)    |
|  step  |   oirf       Lower       Upper    |   oirf       Lower       Upper    |
|--------+-----------------------------------+-----------------------------------|
|0       | 0           0           0         | .84874      .796082     .901397   |
|1       | .045567     -.034257    .125391   | .179558     .098727     .260389   |
|2       | .015859     -.012261    .043979   | .053611     .02135      .085873   |
|3       | .005042     -.005332    .015417   | .01678      .001498     .032061   |
|4       | .001589     -.002218    .005396   | .005279     -.001314    .011872   |
|5       | .0005       -.000874    .001874   | .001662     -.000994    .004317   |
+-------------------------------------------------------------------------------+
95% lower and upper bounds reported
(1) irfname = order1, impulse = y, and response = y
(2) irfname = order1, impulse = y, and response = x
(3) irfname = order1, impulse = x, and response = y
(4) irfname = order1, impulse = x, and response = x


. irf graph oirf, impulse(y x) response(y x)
```



Due to the graph above, x causes x, x causes y, and y causes y present a positive impact while y causes x presents a negative impact. The magnitude is uncertain because the size of the grey area is quite difficult to determine just by looking. The largest grey area represents the largest magnitude.

Only from the picture, all cases seem to have a stable system. The effect seems to die down when time passes.

Unfortunately, some part of the grey area may cross the zero-horizontal line. This shows that they may be insignificant.

```
. irf table coirf, impulse(y x) response(y x)
```

Results from order1

	(1)	(1)	(1)	(2)	(2)	(2)
step	coirf	Lower	Upper	coirf	Lower	Upper
0	.991397	.929889	1.0529	-.519194	-.60033	-.438057
1	1.09882	.988608	1.20903	-.2891	-.415583	-.162616
2	1.12583	.997531	1.25413	-.203589	-.345407	-.06177
3	1.13411	.99822	1.26999	-.176236	-.325307	-.027165
4	1.13671	.997873	1.27554	-.167611	-.319701	-.01552
5	1.13752	.997614	1.27743	-.164895	-.318155	-.011636

	(3)	(3)	(3)	(4)	(4)	(4)
step	coirf	Lower	Upper	coirf	Lower	Upper
0	0	0	0	.84874	.796082	.901397
1	.045567	-.034257	.125391	1.0283	.925928	1.13067
2	.061425	-.04613	.168981	1.08191	.957756	1.20606
3	.066468	-.051223	.184159	1.09869	.965427	1.23195
4	.068057	-.053323	.189437	1.10397	.967032	1.2409
5	.068557	-.054144	.191258	1.10563	.967297	1.24396

95% lower and upper bounds reported

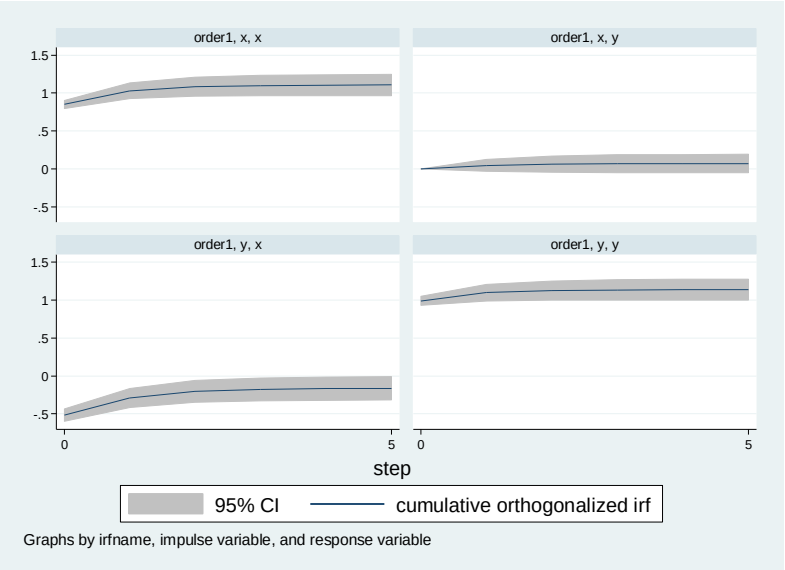
(1) irfname = order1, impulse = y, and response = y

(2) irfname = order1, impulse = y, and response = x

(3) irfname = order1, impulse = x, and response = y

(4) irfname = order1, impulse = x, and response = x

```
. irf graph coirf, impulse(y x) response(y x)
```



Due to the graph above, all cases present a positive impact. The magnitude is uncertain because the size of the grey area is quite difficult to determine just by looking. To me, the largest magnitude is y causes x. The largest grey area represents the largest magnitude.

Only from the picture, all cases seem to have a stable system. The effect seems to die down when time passes.

Unfortunately, some part of the grey area may cross the zero-horizontal line. This shows that they may be insignificant.

According to the IRF analysis, yt seems to have more impact on xt.

(d) Perform Forecast error variance decomposition (fevd) and determine variable that has more impact on each endogenous variable. (3 points)

```
. *FEVD


. irf table fevd, impulse(y x) response(y x)

                                    Results from order1

+-------------------------------------------------------------------------------+
|        |    (1)         (1)         (1)    |    (2)         (2)         (2)    |
|  step  |   fevd       Lower       Upper    |   fevd       Lower       Upper    |
|--------+-----------------------------------+-----------------------------------|
|0       | 0           0           0         | 0           0           0         |
|1       | 1           1           1         | .272307     .205671     .338942   |
|2       | .997916     .990625     1.00521   | .299976     .235166     .364785   |
|3       | .997666     .98951      1.00582   | .303898     .239641     .368155   |
```

4	.997641	.989385	1.0059	.304298	.240105	.368492
5	.997638	.989371	1.00591	.304338	.240151	.368525

	(3)	(3)	(3)	(4)	(4)	(4)
step	fevd	Lower	Upper	fevd	Lower	Upper
0	0	0	0	0	0	0
1	0	0	0	.727693	.661058	.794329
2	.002084	-.005208	.009375	.700024	.635215	.764834
3	.002334	-.005823	.01049	.696102	.631845	.760359
4	.002359	-.005897	.010615	.695702	.631508	.759895
5	.002362	-.005906	.010629	.695662	.631475	.759849

95% lower and upper bounds reported

(1) irfname = order1, impulse = y, and response = y

(2) irfname = order1, impulse = y, and response = x

(3) irfname = order1, impulse = x, and response = y

(4) irfname = order1, impulse = x, and response = x

According to the FEVD, y has more impact on x as can be seen in the FEVD value above.

(e) Determine whether changing Cholesky order from – “yt xt” to “xt yt” will change the results of irf, oirf, coirf, and fevd. Why? or why not? (2 points)

Yes, it would change the result of irf, oirf, coirf and fevd. Different Cholesky order would result in different contemporaneous effects of each variables. Theoretical framework should be used for setting the order.

Moreover, as variance decomposition contains identification problem, the Cholesky order form should be applied to set up the variables order.